

25 Jahre Literarische und Dokumentarische Datenverarbeitung an der Universität Tübingen - ein kurzer Rückblick

Ott, Wilhelm

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Ott, W. (1996). 25 Jahre Literarische und Dokumentarische Datenverarbeitung an der Universität Tübingen - ein kurzer Rückblick. *Historical Social Research*, 21(2), 147-152. <https://doi.org/10.12759/hsr.21.1996.2.147-152>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

bezogenheit von Einzeltexten und erlaubt deren Rekontextualisierung, sie macht das personelle Beziehungsnetz der Zeitung zu Publizisten, Literaten und Politikern sichtbar und öffnet den medienwissenschaftlichen Untersuchungsweg für pressehistorische und strukturelle Forschungsfragen.

Ein Überblick über die Grundlagen-Probleme der Feuilletonforschung wie der Quellenerschließung verbunden mit einer vergleichenden Studie zur Rhetorik zweier Feuilleton-Jahrgänge der *Frankfurter Zeitung* (1919 und 1929) erscheint 1995 (»Das Feuilleton der *Frankfurter Zeitung*. Zur Grundlegung einer rhetorischen Medienforschung«, Max Niemeyer Verlag, Tübingen).

25 Jahre Literarische und Dokumentarische Datenverarbeitung an der Universität Tübingen - ein kurzer Rückblick -

*Wilhelm Ott (ZDV, Tübingen)**

1. Ein runder Geburtstag

Vor 25 Jahren, am 1. September 1970, hat die Universität Tübingen die Förderung der Anwendung der EDV in den Geisteswissenschaften explizit in den Aufgabenkatalog ihrer zentralen Einrichtung »Zentrum für Datenverarbeitung« geschrieben, indem sie dort eine eigene Abteilung mit dem Namen »Literarische und Dokumentarische Datenverarbeitung« eingerichtet hat.

»Aufgabe dieser Abteilung ist die Entwicklung und Betreuung von Methoden und Programmen zur Verarbeitung von Textdaten aller Art. Dadurch soll vor allem den geisteswissenschaftlichen Fachbereichen der Universität Zugang zu dem für sie relativ neuen Hilfsmittel EDV ermöglicht werden. Außerdem sollen damit die Voraussetzungen für die Koordination von Vorhaben geschaffen werden, die mit der Übernahme von Dokumentationsaufgaben auf EDV-Anlagen zusammenhängen und die auch in den übrigen Fachbereichen anfallen.« (»Mitteilungen an unsere Benutzer« des ZDV vom September 1970, S. 5).

Universitätsleitung und Leitung des Rechenzentrums haben damals Weitsicht bewiesen. Denn: Dienstleistung für Geisteswissenschaftler als Aufgabe des zentralen Rechenzentrum - das war damals keineswegs selbstverständlich, nicht nur in Deutschland. Die Rechenzentren waren zwar zentrale Einrichtungen

* Protokoll des 65. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen am 18. November 1995.

gen der Universitäten, aber fest in der Hand der Naturwissenschaften. Was Rechner und Peripherie angeht, waren sie so gut wie ausschließlich ausgelegt auf »number crunching«. Dem entsprach auch die personelle Ausstattung und folglich die Möglichkeit, dort methodischen Rat und Unterstützung zu erhalten. Außer für die Unterstützung naturwissenschaftlicher Aufgabenstellungen war man noch einigermaßen vorbereitet auf Aufgaben aus dem Bereich der mathematischen Statistik. Umgang mit Texten beschränkte sich in der Regel auf Verwaltung und Korrektur der Quellen von FORTRAN-Programmen (die ja noch über Lochkarten eingegeben wurden).

Seither hat sich die Situation grundlegend gewandelt. Schon lange steht das eigentliche Rechnen, der Umgang mit Zahlen, nicht mehr im Mittelpunkt des Bewußtseins, wenn in der Öffentlichkeit über Computer und deren Anwendung gesprochen wird. Die Diskussion wird beherrscht von Begriffen wie Datenauto-
bahn, weltweite Vernetzung, Surfen im Internet, electronic mail, Multimedia, Computer-Spiele oder Tele-Heimarbeit als Lösung für die Verkehrsprobleme auf den deutschen Straßen.

Für die tägliche Arbeit kommt dazu die sog. Textverarbeitung, die auch in der Universität zu den häufigsten Computer-Anwendungen gehören dürfte (Stichworte: Sekretariatsarbeiten oder Tippen einer Seminararbeit oder Vorbereitung einer Druckvorlage für eine Publikation).

Dies ist der Hintergrund, vor dem wir auf 25 Jahre Literarische und Dokumentarische Datenverarbeitung zurückblicken. Die »Nichtnumerik« hat also den numerischen Computer-Anwendungen längst den Rang abgelaufen.

Doch diese Art von Nichtnumerik war wohl nicht das, was die Dekane von vier Fachbereichen am 25. Mai 1970 in einem Brief an das Akademische Rektorat meinten, wenn sie die Errichtung einer Abteilung »Nichtnumerik« forderten. Der Name war entlehnt von einer gleichnamigen Abteilung des 1961 gegründeten Deutschen Rechenzentrums in Darmstadt. Diese umfaßte die beiden Gruppen »Informationswissenschaften« und »Geisteswissenschaften« und hatte u.a. ein Unterprogrammpaket zur Zeichen- und Stringverarbeitung in FORTRAN entwickelt, auf das auch wir uns zunächst stützten.

Unter dem Oberbegriff »nichtnumerische Computer-Anwendungen« wurden inzwischen aber auch Aufgaben bearbeitet, die mit unserem Arbeitsgebiet nichts zu tun hatten; in Tübingen gehörten dazu das automatische Beweisen oder die Lösung von Schachproblemen.

International war inzwischen der Begriff »Linguistic Computing« in Gebrauch, der auch in der deutschen Übersetzung »Linguistische Datenverarbeitung« zunächst alles umfassen sollte, was mit der Verarbeitung von Sprache und schriftlich fixierten sprachlichen Äußerungen zu tun hatte. Er begegnet uns im Namen des 1973 mit einer Münchener Arbeitstagung initiierten und 1975 gegründeten Vereins »LDV-Fittings« (der heutigen GLDV, »Gesellschaft für Linguistische Datenverarbeitung«) - in der sich jedoch die computer-anwendenden Literaturwissenschaften nicht so recht heimisch fühlen.

Der Name »LDV« erschien angesichts der »fortschreitenden Etablierung des neukonzipierten Faches Linguistik an Universitäten und Hochschulen« für die in Tübingen zu etablierende Einrichtung jedoch nicht mehr brauchbar. Linguistik wurde inzwischen gesehen als »eine empirisch-theoretische Wissenschaft, deren Forschungsbereich die verbale Kommunikation in natürlichen Sprachen ist« (Zitat aus dem »Lexikon der Germanistischen Linguistik« von 1973).

Auch die 1973 in England gegründete »Association for Literary and Linguistic Computing« hat sich einen Doppelnamen gegeben, nach dem Vorbild des von Roy Wisbey geleiteten »Literary and Linguistic Computing Centre« der Universität Cambridge, das 1964 noch unter dem Namen »Linguistic Computing Centre« gegründet worden war und 1966 umbenannt wurde.

Für den ersten Bestandteil unseres Namens gab es also Vorbilder. Dem zweiten Arbeitsgebiet, das in der Aufgabenbeschreibung von 1970 genannt war und Textdatenverarbeitung auch außerhalb der eigentlichen Literaturwissenschaften einschließt (z.B. das Erarbeiten von Lexika, Wörterbüchern, Bibliographien und anderen Verzeichnissen), sollte der Namensbestandteil »dokumentarisch« Rechnung tragen.

So viel also zur »Taufe« unseres Geburtstagskindes.

2. »An ihren Früchten sollt ihr sie erkennen«

Statt einer Liste weiterer Daten und Fakten soll eine Art »Fotoalbum« einen Überblick über weitere Stationen der Entwicklung der letzten 25 Jahre bieten. Die Rolle der Bilder von bemerkenswerten Ereignissen soll dabei eine Auswahl von (in diesem Protokoll nicht reproduzierten) Titelblättern und Beispielseiten von Publikationen aus Projekten übernehmen, an denen die Abteilung LDDV beratend oder betreuend beteiligt war oder die mit Hilfe der hier entwickelten Werkzeuge durch weitgehend selbständige Bearbeiter realisiert wurde.

Das erste Beispiel stammt aus der klassischen Philologie. 1966 hatte ich am Deutschen Rechenzentrum in Darmstadt begonnen, epische Dichtung der klassischen römischen Literatur metrisch zu analysieren und nach dem Vorbild der Anhänge von Eduard Nordens Kommentar zum 6. Buch von Vergils Aeneis Material zum lateinischen Hexameter zusammenzustellen, das dank Computer-Hilfe jetzt erstmals vollständig und nachprüfbar sein konnte. Die »Metrischen Analysen zur Ars Poetica des Horaz«, die 1970 bei Kümmerle in Göttingen publiziert wurden, sind gleichzeitig das erste Buch, das mit dem Vorläufer des heutigen TUSTEP-Satzprogramms direkt von den bei der Analyse gewonnenen Dateien automatisch (und damit ohne das Risiko von Druckfehlern) im Lichtsatz gesetzt und umbrochen wurde.

Das nächste Beispiel weist schon über Tübingen hinaus: Harald Zimmermann von der Universität Saarbrücken hatte nachgefragt, ob wir helfen könnten, die Druckvorlage für den Index zu Georg Trakls Dichtungen (erschienen 1971 in Frankfurt bei Athenäum) ähnlich fehlerfrei und ästhetisch befriedigend

zu erstellen. Dies war die erste Bewährungsprobe eines ursprünglich für einen anderen Zweck erstellten Programms - das freilich im Hinblick auf weitere Verwendbarkeit schon recht modular angelegt war.

Denn es war von vornherein beabsichtigt, mit Hilfe dieser Bausteine später die Druckvorlagen für ein DFG-gefördertes Projekt aus dem Beuroner Vetus-Latina-Institut zu erstellen, für die 1977 in 5 Bänden bei Frommann-Holzboog in Stuttgart erschienene Vulgata-Konkordanz.

Die Arbeiten daran hatten Ende der 60er Jahre begonnen. An diesem Projekt haben wir viel für unsere spätere Arbeit gelernt. Dazu gehörte u.a. die Erkenntnis, daß die sonst im Umlauf befindlichen »Basis-Programme« für philologischen Umgang mit Texten zu viele Grundfunktionen in sich vereinigten. So z.B. das in England 1967 entstandene Programm für »word count and concordance generation on Atlas« COCOA, aus dem 1979/80 das weltweit verbreitete Oxford Concordance Program (OCP) hervorgegangen ist. Ein ähnliches Programm war das 1973/74 in Regensburg entwickelte »Basisprogramm für die Textverarbeitung« COBAPH, in dessen Mittelpunkt die Erstellung von Frequenzwörterbüchern, Wortindizes und Konkordanzen stand. Um wirklich flexibel auf unterschiedliche Anforderungen reagieren zu können, werden Bausteine gebraucht, die relativ elementare Grundoperationen einzeln bereitstellen, für die Index- und Konkordanzarbeit also z.B. je einen eigenen Baustein für das Zerlegen des Textes, für das Hinzufügen von Sortierschlüsseln, für die Sortierung selbst, ggf. für das Einmischen anderer Daten, für das Zusammenfassen von Zusammengehörigem nach der Sortierung und schließlich für das Formatieren und Drucken des fertigen Index. Die konsequente Modularität von TU-STEP kommt aus dieser Einsicht.

Das viertes Beispiel kommt aus dem Deutschen Seminar der Universität Tübingen: Paul Sapplers Edition von Heinrich Kaufringers Werken und die dazugehörigen Indizes (Tübingen: Niemeyer 1972 und 1974). In der Festschrift zum 125-jährigen Jubiläum des Max Niemeyer Verlags (Tübingen 1995) ist diese Edition angeführt »als erstes in Deutschland im Computer gesetztes und umbrochenes Buch, bei dem auch die Fußnoten automatisch umbrochen wurden«. Der eigentliche Zweck des Computer-Einsatzes war aber die Erarbeitung des Index-Bandes, der das Wortmaterial dieser um 1400 entstandenen Dichtung nach allen Regeln der philologischen Kunst darbieten sollte (und sich darin von vielen damals publizierten EDV-Indizes unterscheidet). Der Satz der Edition fiel sozusagen als »Nebenprodukt« ab.

Hier möchte ich die chronologische Darstellung verlassen. Der Einschnitt an dieser Stelle erscheint passend: Denn Sapplers Kaufringer-Index markiert die Stelle, an der die Nutzer des Werkzeuges EDV selbständig werden, mit der Zeit immer unabhängiger von der konkreten Programmier-Unterstützung durch Mitarbeiter des ZDV. Für Paul Sappler war dies noch relativ mühsam: es stand - außer dem Satzprogramm und Basis-Routinen für die Register-Arbeit - an Standard-Bausteinen anfangs noch nicht viel mehr zur Verfügung als das schon

erwähnte Paket von Unterprogrammen zu FORTRAN, die ein einigermaßen komfortables Umgehen mit Textdaten in dieser Programmierumgebung erlaubten. Programmieren mußte er vieles noch selber.

Die inhaltliche Verschiedenheit der betreuten Projekte und der wachsende Bedarf an Unterstützung weiterer Projekte machte es notwendig, nach neuen Wegen zu suchen. Wir mußten auch Wissenschaftler ohne FORTRAN-Kenntnisse und ohne Ambitionen in dieser Richtung unterstützen können, ohne selbst auf ein Einzelprojekt zugeschnittene Programme schreiben zu müssen, aber auch ohne gleichzeitig die notwendige Flexibilität zu verlieren.

Nur so konnten wir auch den Auftrag vom 1.9.1970 eigentlich erfüllen. In dem oben zitierten Brief vom Mai 1970 an das Akademische Rektorat hatten die Unterzeichner »die *Programmierung und Ausführung* nichtnumerischer Arbeitsaufträge der Geistes- und Kulturwissenschaftlichen Fächer« als Dienstleistung erwartet. Bei der Definition der Aufgaben war dann aber - in weiser Voraussicht - die »*Entwicklung und Betreuung von Methoden und Programmen zur Verarbeitung von Textdaten aller Art*« an die Stelle der Ausführung von Einzelaufträgen gesetzt worden. Die dazu notwendige Erfahrung mußte freilich zunächst aus der Programmierarbeit für Einzelprojekte und aus der engen Kooperation mit diesen gewonnen werden.

Es vergingen sieben Jahre, bis wir der übertragenen Aufgabe so weit nachgekommen waren, daß wir auch die Ausbildung komplett umstellen konnten von Kursen in FORTRAN und in der Anwendung der Unterprogramme zur Zeichen- und Stringverarbeitung und sie im Jahr 1977 ersetzen konnten durch die Einführung in die Benutzung der neuen, keine FORTRAN-Kenntnisse mehr voraussetzenden Grundbausteine. 1978 hat dieses System von Programmen zur »Verarbeitung von Textdaten aller Art« den Namen TUSTEP erhalten.

Kehren wir zurück zu unserem Fotoalbum. Die im Kolloquium vorgestellten Beispiele können hier nur summarisch aufgeführt werden. Sie lassen sich in folgende inhaltliche Gruppen zusammenfassen:

1. Editionen: Die Liste der mit TUSTEP erstellten und/oder gesetzten Editionen umfaßt inzwischen mehr als 150 Bände mit zusammen mehr als 97.000 Druckseiten, darunter als international bekannteste Hans-Walter Gablers Edition des Ulysses von James Joyce (New York: Garland 1984); auch viele Bände aus der Reihe des Frankfurter Deutschen Klassiker Verlags gehören dazu.
2. Konkordanzen, Indizes: neben der Vulgata-Konkordanz seien die hebräische »Konkordanz zur Hekhalot-Literatur« (Tübingen: Mohr 1988) und die »Sanskrit Indices and Text of the Brahmapurāṇa« (Wiesbaden: Harrassowitz 1987) angeführt.
3. Bibliographien, Repertorien, Kataloge und ähnliche Verzeichnisse. Die bekanntesten Vertreter sind die Periodica »Germanistik« (seit 1977 mit TUSTEP erstellt) und »Romanische Bibliographie« (seit dem Berichtsjahr 1975/76) des Tübinger Niemeyer-Verlags oder die allein inzwischen mehr

als 7000 A4-Seiten (zweispaltig) umfassenden Neuerwerbungslisten der Universitätsbibliothek. Insgesamt umfaßt meine Sammlung an Belegexemplaren derzeit mehr als 190 Bände mit mehr als 71.000 Druckseiten, darunter auch die jährlich erscheinenden und am Ende jeder Legislaturperiode kumulierten umfangreichen »Register zu den Verhandlungen des Deutschen Bundestages und des Bundesrats«, die aus der Bonner Datenbank extrahiert und seit 1976 mit TUSTEP gesetzt werden.

4. Lexika und Wörterbücher, als Vertreter seien genannt »Deutsches Wörterbuch« von Hermann Paul (Tübingen- Niemeyer 1992) und »Neuhochdeutscher Index zum mittelhochdeutschen Wortschatz« (Stuttgart Hirzel 1990).
5. Sprachliche und inhaltliche Analyse von Texten: Stellvertretend für viele sollen daher am Ende zwei offensichtliche Beispiele aus diesem Bereich stehen, nämlich Winfried Baders 1991 veröffentlichte Dissertation, bei der der EDV-Bezug schon aus dem Untertitel »Computerlinguistische Interpretation des Textes Ri 13-16« hervorgeht, und Jörg Rüpkens 1995 publizierte Habilitationsschrift aus der Fakultät für Kulturwissenschaften über »die Geschichte der Repräsentation und religiösen Qualifikation von Zeit in Rom«.

Wer öfter an diesem Kolloquium teilnimmt, wird viele der erwähnten Projekte wiederkennen und sich erinnern an die Vielfalt unterschiedlichster Aufgabenstellungen aus vielen geisteswissenschaftlichen Disziplinen und an die inhaltlichen Probleme, zu deren Lösung die EDV nicht unwesentlich beigetragen hat.

3. Das organisatorische Umfeld

Das Blättern im Fotoalbum sollte exemplarisch einige der Früchte der Tatsache zeigen, daß es für die Unterstützung der Geisteswissenschaften durch EDV in Tübingen eine zentrale Anlaufstelle gibt, in der die Erfahrungen aus vielen Einzelprojekten gesammelt und weitergegeben und in geeignete Werkzeuge umgesetzt werden konnten und können. Es konnte vielleicht auch sichtbar machen, welche Bedeutung das in Zusammenarbeit mit diesen Projekten entwickelte und vervollkommnete Werkzeug für die geisteswissenschaftliche Forschung inzwischen erlangt hat

Ich nehme an, daß ich mich an dieser Stelle zum Sprecher für Sie alle machen darf, wenn ich den an der Universität dafür Verantwortlichen unseren Dank ausspreche, nicht nur für die Weitsicht, die sie vor 25 Jahren bewiesen haben, sondern auch für die Kontinuität, mit der sie die Fortführung dieser Arbeiten auch in Zeiten ermöglicht haben, in denen die Mittel knapp wurden oder sich andere Aufgaben der EDV-Versorgung in den Vordergrund drängten.

Ein besonderer Dank gebührt auch Land Baden-Württemberg, das in den Jahren 1985-1989 im Forschungsförderungsprogramm des Landes mit großzügig bemessenen Mitteln die Arbeit an den hier entwickelten Methoden und Werkzeugen unterstützt und für dringend notwendige Verbesserung der Infrastruktur gesorgt hat. Nach Ablauf der Schwerpunkt-Förderungsperiode hat es